

1/10

Figure 1: Top blastp results for INSP161 polypeptide sequence (SEQ ID NO: 18) against the NCBI-nr database.

BLASTP 2.2.2 [Jan-08-2002]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query= insp161.pep
(470 letters)

Database: All non-redundant GenBank CDS
translations+PDB+SwissProt+PIR+PRF
1,523,012 sequences; 490,363,361 total letters

Searching.....done

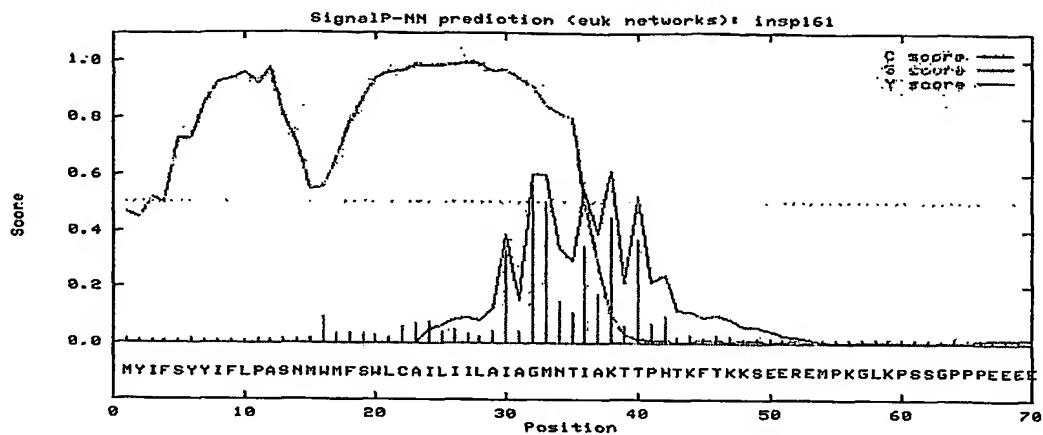
	Score	E
Sequences producing significant alignments:		(bits) Value
ref XP_067228.5 similar to otolin-1 [Homo sapiens]	907	0.0
ref XP_143327.1 similar to INNER EAR-SPECIFIC COLLAGEN PRECURSO...	622	e-177
ref XP_227256.1 similar to INNER EAR-SPECIFIC COLLAGEN PRECURSO...	530	e-149
sp P83371 OTO1_ONCKE Otolin-1 precursor >gi 18496364 dbj BAB8456...	365	e-100
sp P98085 COLE_LEPMA INNER EAR-SPECIFIC COLLAGEN PRECURSOR (SACC...	351	2e-95
pir A55797 collagen precursor, saccule-specific - bluegill >gi ...	325	1e-87
ref XP_224253.1 similar to hypothetical protein FLJ31208 [Homo ...	257	4e-67
ref NP_848635.1 hypothetical protein MGC48915 [Homo sapiens] >g...	249	6e-65
ref XP_290602.1 similar to Adiponectin precursor (30 kDa adipoc...	249	1e-64
ref NP_777059.1 collagen, type X, alpha 1 (Schmid metaphyseal c...	246	6e-64

2/10

Figure 2: Signal peptide prediction (SignalP V2.0) output for INSP161 polypeptide sequence (SEQ ID NO: 18).

>insp161

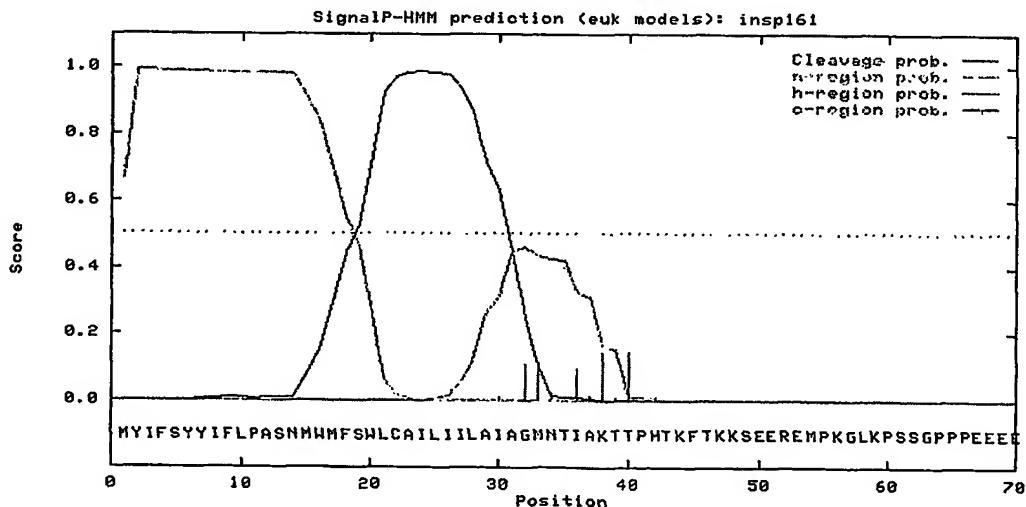
SignalP-NN results:



data:

```
>insp161
# Measure  Position  Value  Cutoff  signal peptide?
max. C      32      0.564  0.33   YES
max. Y      38      0.613  0.32   YES
max. S      28      0.992  0.82   YES
mean S  1-37  0.803  0.47   YES
# Most likely cleavage site between pos. 37 and 38: TIA-KT
```

SignalP-HMM result:



data

```
>insp161
Prediction: Signal peptide
Signal peptide probability: 0.667
Signal anchor probability: 0.326
Max cleavage site probability: 0.146 between pos. 37 and 38
```

3/10

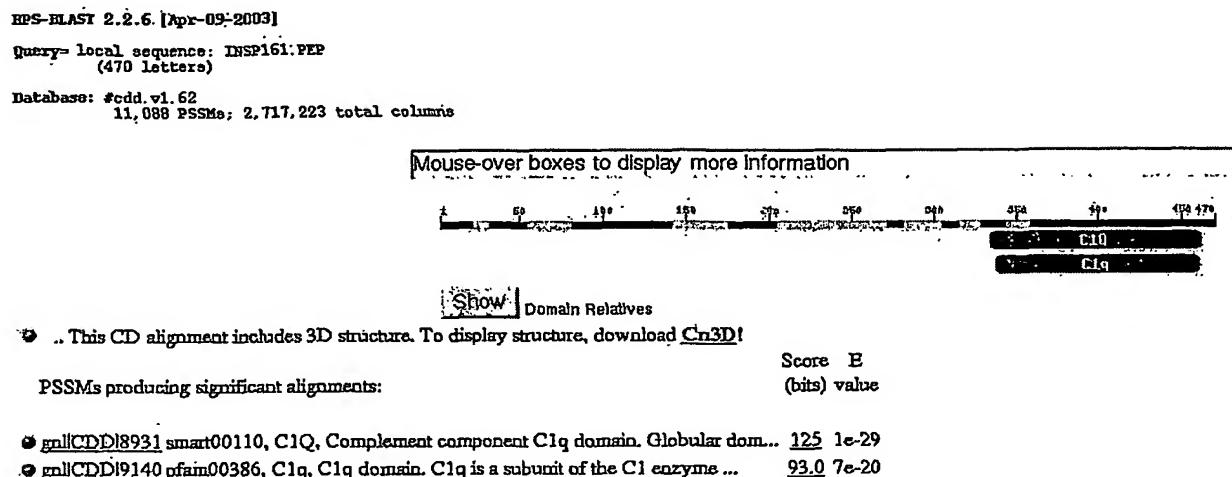
Figure 3: NCBI CDD output for INSP161

Figure 4: INSP161 nucleotide and amino acid sequence

1 gcataatcatt cattctttag attcatatac atatatattt cttttatatg agtataatgtt
 61 tatatgactt agtctttca ttggacctc acataatatg tatatatattt cctattatat
 Signal peptide
 m y i f s y y

121 ctttcttcca gcttcaaata tggatgtt ttctggctt tggctatattt taattatattt
 i f l p a s n m w m f s w l c a i l i i

181 ggctattgct ggtatgaaca caatagcaaa gaccacacca cataccaaat ttacgaagaa
 l a i a g m n t i a k t t p h t k f t k

241 atctgagggaa agagagatgc caaagggtct aaagccatcc agtggcccac ctccagaaga
 k s e e r e m p k g l k p s s g p p p e

301 agaagaaacc ctcttacacag aaatggctga aatggcagaa ccaattacca aaccctcgcc
 e e e t l f t e m a e m a e p i t k p s

361 ctggattct gtcttggca ctgccactct ctctccctt gaaaacttca ctcttgacc
 a l d s v f g t a t l s p f e n f t l d

421 agctgatttc ttttgaatt gttgtgattt ttgttcaccc taccggggc agaaaggaga
 p a d f f l n c c d c c s p v p g q k g

481 acctggagag actggacagc caggtcctaa aggagaggct gaaatttgg ggatcccagg
 e p g e t g q p g p k g e a g n l g i p

541 gccaccagga gttgtgggc cccaggccc tagaggctac aaaggagaga aaggtgaacc
 g p p g v v g p q g p r g y k g e k g e

601 tggccctaag ggagataaag gaaacattgg tttgggagga gtgaaaggac aaaaaggctc
 p g p k g d k g n i g l g g v k g q k g

661 caagggagac acatgtggga attgtaccaa aggagaaaaa ggagaccaag gggctatgg
 s k g d t c g n c t k g e k g d q g a m

721 ctcacctggc ctgcacggag ggcctggcgc caagggagag aagggggaga tgggggagaa
 g s p g l h g g p g a k g e k g e m g e

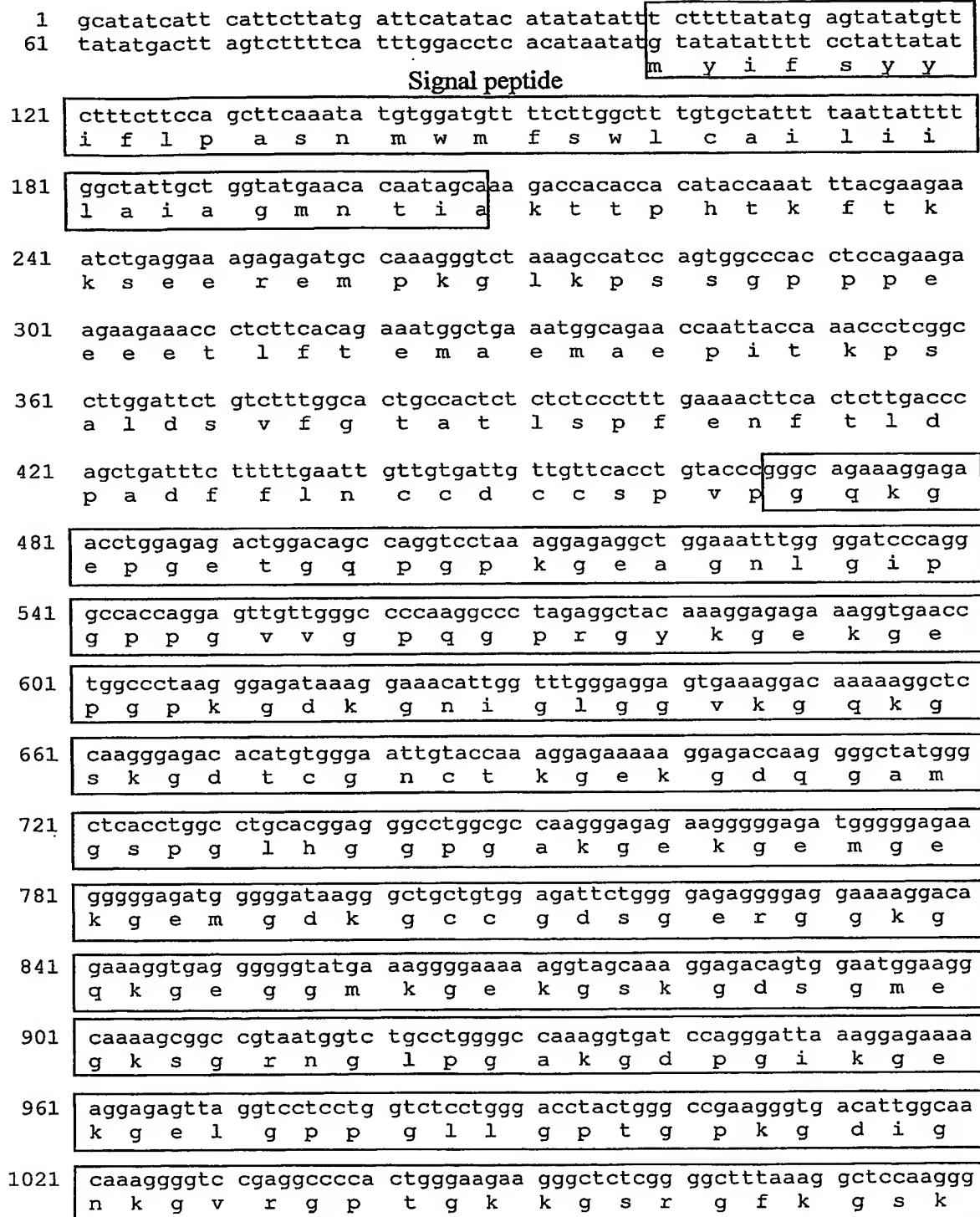
781 gggggagatg gggataagg gctgctgtgg agattctggg gagaggggag gaaaaggaca
 k g e m g d k g c c g d s g e r g g k g

841 gaaaggtag ggggtatga aagggaaaaa agtagcaaa ggagacagtg gaatggaggg
 q k g e g g m k g e k g s k g d s g m e

901 caaaagcggc cgtaatggtc tgcctgggc caaaggatg ccaggattaaaggaaaa
 g k s g r n g l p g a k g d p g i k g e

961 aggagagtta ggtccctctg gtctcctggg acctactggg ccgaagggtg acattggcaa
 k g e l g p p g l l g p t g p k g d i g

1021 caaagggtc cgaggccccca ctggaaagaa gggctctcg ggcttaaag gctccaagg
 n k g v r g p t g k k g s r g f k g s k


 The diagram illustrates the structure of the INSP161 protein. It shows the nucleotide sequence with amino acid translations. A signal peptide is indicated at the top. Below the sequence, 12 vertical boxes are stacked, each representing a collagen domain. The domains are labeled on the left with their start positions (121, 181, 241, 301, 361, 421, 481, 541, 601, 661, 721, 781, 841, 901, 961, 1021). The domains are labeled 'Collagen domains' on the right, with an upward-pointing arrow above the first domain and a downward-pointing arrow below the last domain. The C-terminal region of the protein follows the last domain.

5/10

1081 tgagtggct agagtcccc ggtcgcttt cagcgctgg ttgtcaaagc catttcctcc
 g e l a r v p r s a f s a g l s k p f p

1141 tcctaacatc cccatcaa at ttgaaaagat tctctataat gaccaaggga attacagtcc
 p p n i p i k f e k i l y n d q g n y s

1201 tgtcactggg aagtttaact gctctattcc tgggacata at gttttttcct accatattac
 p v t g k f n c s i p g t y v f s y h i

1261 ggtgaggggg cgacctgctc gaatcagttct ggtggccag aataagaagc agttcaagtc
 t v r q r p a r i s l v a q n k k q f k

1321 cagagaaact ctctatggtc aggaaataga ccaggcctct ctcctcgta tcttggaaatt
 s r e t l y g q e i d q a s l l v i l k

1381 aagtgcagga gaccaagtct ggcttgaggt gtc当地atgggg tggatgggg tggatgtcag
 l s a g d q v w l e v s k d w n g v y v

1441 tgctgaggat gacagcattt ttactgggtt cttttgtac ccagaggaaa cttctggaaat
 s a e d d s i f t g f l l y p e e t s g

1501 ttcaccataa atttgtgtcc tgaatcctgt agtttagatt cagtggata agtcagttaa
 i s p

1561 cacagagtag tgctattaaa aaataacttc catttttca agattat

C1q domain

Figure 5

A)

1 SSGPPPEEEE TLFTEMAEMA EPITKPSALD SVFGTATLSP FENFTLDPAD

51 FFLNCCDCCS PVPGQKGEPEG ETGQPGPKGE AGNLGIPGPP GVVGPGPRG
Collagen region

101 YKGEKGEPEGP KGDKGNIGLG GVKGQKGSKG DTGNCTKGE KGDQGAMGSP
Collagen region

151 GLHGGPGAKG EKGEMGEKGE MGDKGCCGDS GERGGKGQKG EGGMKGEKGS
Collagen region

201 KGDSGMEGKS GRNGLPGAKG DPGIKGEKGE LGPPGILLGPT GPKGDIGNKG
Collagen region

251 VRGPTGKKGS RGFKGSKGET ARVPRSAFSA GLSKPFPPPN IPIKFEKILY
Clq domain

301 NDQGNYSPTV GKFNCSIPGT YVFSYHITVR GRPARISLVA QNKQFKSRE
Clq domain

351 TLYGQEIDQA SLIVIILKISA GDQVWLEVSK DWNGVYVSAE DDSIFTGFLL

401 YPEETSGISP

B)

1 SSGPPPEEEE TLFTEMAEMA EPITKPSALD SVFGTATLSP FENFTLDPAD

51 FFLNCCDCCS PVPGQKGEPEG ETGQPGPKGE AGNLGIPGPP GVVGPGPRG
Collagen region

101 YKGEKGEPEGP KGDKGNIGLG GVKGQKGSKG DTGNCTKGE KGDQGAMGSP
Collagen region

151 GLHGGPGAKG EKGEMGEKGE MGDKGCCGDS GERGGKGQKG EGGMKGEKGS
Collagen region

201 KGDSGMEGKS GRNGLPGAKG DPGIKGEKGE LGPPGILLGPT GPKGDIGNKG

251 VRGPTGKKGS RGFKG

C)

1 SKGELARVPR SAFAAGLSKP FPPPNIPIKF EKILYNDQGN YSPVTGKFC
Clq domain

51 SIPGTYVFSY HITVRGRPAR ISLVAQNKQ FKSRETLYGQ EIDQASLLVI
Clq domain

101 LKISAGDQVW LEVSKDWNGV YVSAEDDSIF TGFLIYPEET SGISP

7/10

Figure 6

hg15_dna	1697	GCATATCATTCACTTATGATTCAATACATATATTTCTTTATATG 	1746
insp161.cdna	1	GCATATCATTCACTTATGATTCAATACATATATTTCTTTATATG	50
hg15_dna	1747	AGTATATGTTATATGACTTAGTCTTCATTTGGACCTCACATAATATG 	1796
insp161.cdna	51	AGTATATGTTATATGACTTAGTCTTCATTTGGACCTCACATAATATG INSP161-AP1	100
hg15_dna	1797	TATATATTTCTATTATATCTTCTTCAGCTCAAATATGTGGATGTT 	1846
insp161.cdna	101	TATATATTTCTATTATATCTTCTTCAGCTCAAATATGTGGATGTT	150
hg15_dna	1847	TTCTTGGCTTGTGCTATTTAATTATTTGGCTATTGCTGGTATGAACA 	1896
insp161.cdna	151	TTCTTGGCTTGTGCTATTTAATTATTTGGCTATTGCTGGTATGAACA	200
hg15_dna	1897	CAATAGCAAAGACACACCACATACCAAATTACGAAGAAATCTGAGGAA 	1946
insp161.cdna	201	CAATAGCAAAGACACACCACATACCAAATTACGAAGAAATCTGAGGAA	250
hg15_dna	1947	AGAGAGATGCCAAAGGTCTAAAGCCATCCAGTGGCCACCTCCAGAAGA 	1996
insp161.cdna	251	AGAGAGATGCCAAAGGTCTAAAGCCATCCAGTGGCCACCTCCAGAAGA	300
hg15_dna	1997	AGAAGAAACCCCTTTCACAGAAATGGCTGAAATGGCAGAACCAATTACCA 	2046
insp161.cdna	301	AGAAGAAACCCCTTTCACAGAAATGGCTGAAATGGCAGAACCAATTACCA	350
hg15_dna	2047	AACCCTCGGCCTTGGATTCTGTCTTGGCACTGCCACTCTCTCCCTTT 	2096
insp161.cdna	351	AACCCTCGGCCTTGGATTCTGTCTTGGCACTGCCACTCTCTCCCTTT	400
hg15_dna	2097	GAAAACCTCACTCTTGACCCAGCTGATTCTTTGAATTGTTGTGATTG 	2146
insp161.cdna	401	GAAAACCTCACTCTTGACCCAGCTGATTCTTTGAATTGTTGTGATTG	450
hg15_dna	2147	TTGTTCACCTGTACCCGGGCAGAAAGGAGAACCTGGAGAGACTGGACAGC 	2196
insp161.cdna	451	TTGTTCACCTGTACCCGGGCAGAAAGGAGAACCTGGAGAGACTGGACAGC INSP161-AP2	500
hg15_dna	2197	CAGgtatt.....tctagGTCTAAAGGAGAGGCTGGAAATTGGGGAT >>> 1999 >>>	4229
insp161.cdna	501	CAG.....GTCCTAAAGGAGAGGCTGGAAATTGGGGAT	534
hg15_dna	4230	CCCAGGGCCACCAGGAGTTGTGGCCCCAAGGCCCTAGAGGCTACAAAG 	4279
insp161.cdna	535	CCCAGGGCCACCAGGAGTTGTGGCCCCAAGGCCCTAGAGGCTACAAAG INSP161-AP3	584
hg15_dna	4280	GAGAGAAAGgtagg.....ttcagGTAACCTGGCCCTAACGGAGATAA >>> 3765 >>>	8078
insp161.cdna	585	GAGAGAAAG.....GTAACCTGGCCCTAACGGAGATAA	618

9/10

hg15_dna	8879	ATTGGAATGGGTGTATGTCAGTGCTGAGGATGACAGCATTACTGGG	8928
insp161.cdna	1419	ATTGGAATGGGTGTATGTCAGTGCTGAGGATGACAGCATTACTGGG	1468
hg15_dna	8929	TTCCTTTGTACCCAGAGGAACTTCTGGAATTCAACCATAAATTGTGT	8978
insp161.cdna	1469	TTCCTTTGTACCCAGAGGAACTTCTGGAATTCAACCATAAATTGTGT	1518
← INSP161-AP6			
hg15_dna	8979	CCTGAATCCTGTAGTTAGATTCACTGGAAATAAGTCAGTTAACACAGAGT	9028
insp161.cdna	1519	CCTGAATCCTGTAGTTAGATTCACTGGAAATAAGTCAGTTAACACAGAGT	1568
hg15_dna	9029	AGTGCTATTAAAAATAACTTCCATTTCAGATTAT	9067
insp161.cdna	1569	AGTGCTATTAAAAATAACTTCCATTTCAGATTAT	1607

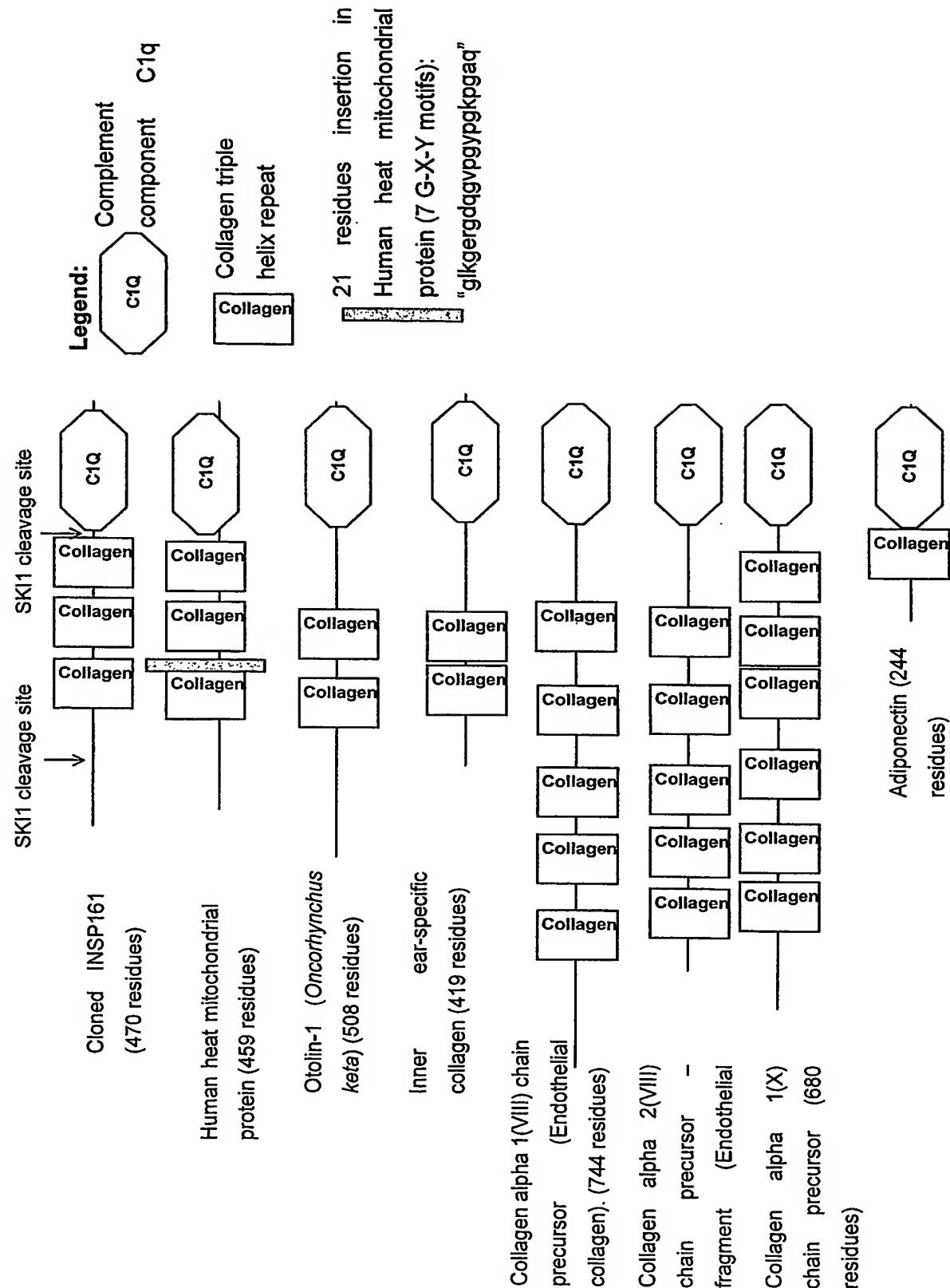


Figure 7